Migrant Boat Mini Challenge Award: Analysis Summary

A Geo-Temporal Analysis of the Migrant Boat Dataset

Benjamin Holland[1]   Lisa Kuchy[2]   Jason Dalton[3]

SPADAC, Inc.

ABSTRACT

The SPADAC team used various visual analytics tools and methods to find geo-temporal patterns of migration from a Caribbean island from 2005-2007. In this paper, we describe the tools and methods used in the analysis. These methods included generating temporal variograms, dendrograms, and proportionally weighted migration maps, using tools such as the *R* statistical software package and *Signature Analyst*™. We found that there is a significant positive space-time correlation with the boat encounters (especially the landings), with a migratory shift further away from the point of departure over time.

KEYWORDS: visual analytics, geo-temporal analysis.

INDEX TERMS: G.1.1 [Numerical Analysis]: Interpolation—Interpolation-Formulas; G.1.2 [Numerical Analysis]: Approximation—Nonlinear Approximation; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Clustering; I.2.1 [Artificial Intelligence]: Applications and Expert Systems—Cartography

1. INTRODUCTION

Geospatial patterns of a particular dataset can be analyzed using various methods and models to understand the full extent of any spatial trends. The analysis becomes "geo-temporal" when temporal as well as spatial information about the data is available. This type of analysis involves finding correlations between the spatial and temporal components, where and when those correlations are found, and to what extent. In this challenge, boat landings and interdictions from a fictional island, Isla del Sueño, were analyzed over a three year period.

2. GEO-TEMPORAL VISUALIZATION

2.1 Overall Geo-temporal Correlation

Initially, we used *Google Earth* to view the overall geo-temporal trends. We were able to dynamically view the boat encounters over different time frames by using the timeline slider bar feature, which enabled us to view animations depicting the migrations. The encounters were separated into interdictions and landings to investigate existing trends within the two groups.
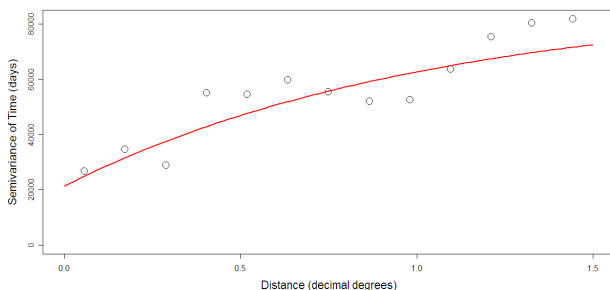
A variogram, which is a commonly used visualization tool in geostatistics, was used to visualize the overall extent of space-time clustering of the boat landings (Figure 1) and interdictions. It shows to what extent a set of points becomes more or less similar in time as they get farther away in space. This can indicate whether there is a significant space-time relationship, and to what extent. We can infer from the variogram in Figure 1 that there is a significant positive space-time relationship. In other words, landings that are close in space tend to be close in time.

The temporal variograms were created using the "GeoR" package in *R*. *R* is a free software environment for statistical computing and graphics originally written by Robert Gentleman and Ross Ihaka at the Statistics Department of the University of Auckland in 1997 and is now supported by a development team.

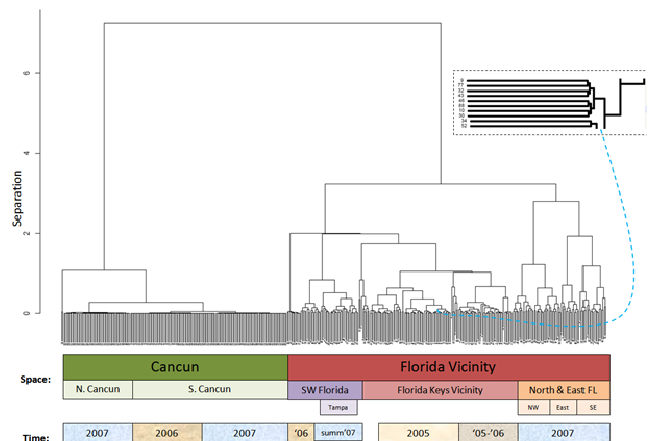### 2.2 Geo-temporal patterns by region

2.2.1 Dendrograms

Figure 2: Temporally sequenced geospatial dendrogram of boat
-    7

Next, we created a temporally sequenced geospatial dendrogram (which shows clusters of points with time stamps in geographic space) to analyze local clusters of landings (Figure 2) and interdictions by year. In contrast with the variogram, which shows overall trends, here we are most interested in local (yearly, and by geographic region) patterns of space-time clustering.

The time stamps are indices which correspond to a given time. In this case, an index of 1 corresponds with 01/01/05 while last index corresponds with 12/31/2007. By examining the time stamps on each set of nodes, we were able to discern whether a noticeably large proportion of values in a given time period are in

[1] ben.holland@spadac.com
[2] lisa.kuchy@spadac.com
[3] jason.dalton@spadac.com

a given geographic region. For example, under the spatial cluster of Northern Cancun, the majority of index values for the individual nodes were greater than 300, which means that most of the landings in that region took place in 2007.

The dendrogram was created in SPADAC's *Signature Analyst* using a hierarchical clustering algorithm. *Signature Analyst*, a commercial software product created by SPADAC Inc. in 2002, is a geospatial predictive modelling tool that utilizes deductive modelling and empirically-based pattern discovery methods for geospatial data. The dendrograms in *Signature Analyst* are typically generated after a predictive model is performed on a dataset. The descriptive color coded "space" and "time" boxes below the dendrogram in Figure 2 were designed specifically for this project to help decipher where and when the geospatial clustering was occurring. Soon after the conclusion of the VAST challenge, it was decided that the boxes will be built into the *Signature Analyst* software for future use since it was highly beneficial for this project.

2.2.2   Proportionally Weighted Migration Maps

Another visual analytics tool that was used to find local geo-temporal patterns in the data is a proportional symbol migration map (Figure 3). The thickness of the arrows is proportional to the number of encounters in a given area. This enabled us to coherently visualize the magnitude of the number of encounters in comparison with where they occurred. Besides answering the questions of "where" and "when", this tool also addresses "how much" or "how many". The map in figure 3 indicates that the majority of landings in 2006 occurred near Cancun and the Florida Keys.

The migration maps were created in *PowerPoint* using *Google Earth* generated maps. Given more time, we would have automated the process of drawing arrows with varying thicknesses using customized scripting in *ArcGIS* rather than manually drawing and calculating/manipulating the relative thickness of the arrows.
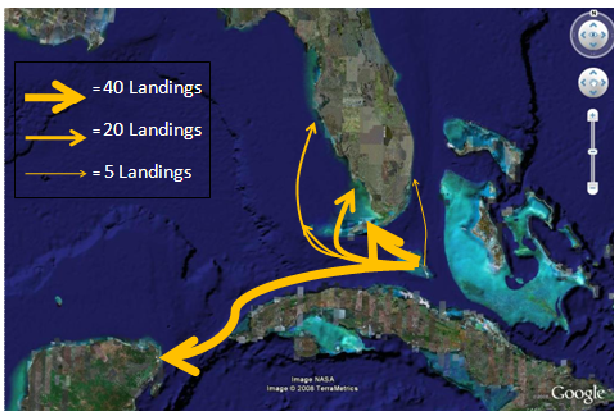

Figure 3: Geographic Patterns of Migration (Landings) from Isla del Sueño, 2006

2.3   Geospatial patterns of successful landings

Figure 4 shows a continuous surface of landing success rates around Florida. The pattern in the map indicates that there is a south to north trend of increasing landing success. It was created by sampling success rates (the number of landings divided by the total number of encounters) across the region using a 40x40 km

moving window and performing inverse distance weighted interpolation (with a power of 1.5) on those sampled points. Inverse distance weighted was chosen over other geospatial
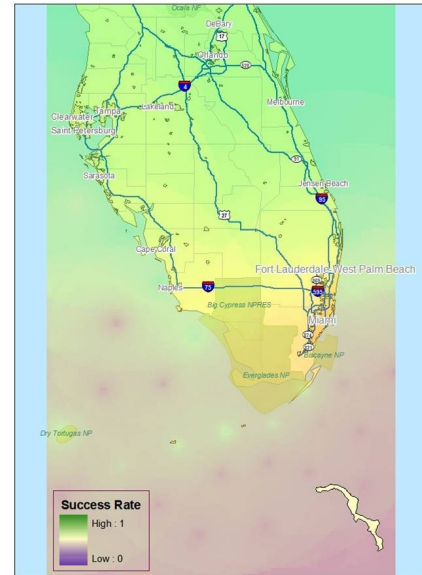

Figure 4: Interpolated map of landing success rate over the three year period

interpolation methods (nearest neighbour and ordinary kriging) because it produced the overall lowest cross validation error.

We performed the interpolation (and created the map) using *ArcGIS 9.2*. *ArcGIS* is a geographic information system software product line, which has capabilities to perform spatial analyses as well as manage and visualize spatial data.

3.   CONCLUSIONS

The SPADAC team used a number of visual analytics tools and various geo-temporal analysis methods to turn the data that was provided into useful knowledge about the scenario.

Given the use of external data, we would have used *Signature Analyst* to create a geospatial assessment of future migrant encounters. As opposed to a density surface which describes only where the past encounters took place, a geospatial assessment exploits information available in the surrounding environment [e.g., ocean currents, levels of commercial and recreational boat traffic, coast guard patrols, and proximity to major cities and borders] by empirically comparing the selected layers to the past events, and projecting the resulting "signature" onto the specified geographical area of interest. The geospatial assessment would show where future encounters are most likely to occur.

4.   REFERENCES

1. "The R Project for Statistical Computing." Department of Science and Mathematics, University of Auckland. June 2008. http://www.r-project.org/
2. "Predictive Analytics." SPADAC Inc. June 2008. http://www.spadac.com
3. "ArcGIS: The Complete Enterprise GIS." ESRI. June 2008. http://www.esri.com